DOCUMENT RESUME

ED 086 003 FL 004 624

AUTHOR Rivers, Wilga M.
TITLE Testing and Student Learning.
PUB DATE 29 Jun 73
NOTE 21p.; Paper presented at the First International Conference of the Association of Teachers of English to Speakers of Other Languages, Dublin, Ireland, June 29, 1973

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Course Content; *Criterion Referenced Tests; Educational Objectives; Independent Study; Individualized Instruction; *Language Instruction; Learning Modalities; Learning Processes; Morphology (Languages); Pacing; Peer Teaching; Phonology; *Second Language Learning; Small Group Instruction; Student Interests; Syntax; Teaching Methods; *Test Construction; *Testing

ABSTRACT

Test design and content and course design and content are mutually influential. In accordance with changing trends in foreign language instruction--individualized learning, independent study, small group instruction, peer teaching--testing should likewise be revamped to reinforce and inform the learning process. Tests must take into account student interests, student ability, performance objectives, preferences among various learning modalities, and willingness to devote time and effort to achieving a particular goal. Criterion referenced tests at their best would incorporate all of these features. (DD)

FILMED FROM BEST AVAILABLE COPY

ED 086003

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TESTING AND STUDENT LEARNING

Wilga M. Rivers
University of Illinois at
Urbana-Champaign

I remember an international conference some six years ago at which a distinguished European scholar whose English was not strong suddenly burst into the discussion of an experimental project with the words: "Why these old tests!" Although he was referring in this case to statistical tests of significance, the words have remained in my mind: Why these old tests! These words are still only too often applicable in practical instructional situations, despite the excellent books and articles of such friends of the profession as John Carroll, Alan Davies, David Harris, Paul Pimsleur, Robert Lado, and Rebecca Valette.

And tests are becoming older. What was new in the sixties does not adequately assess achievement in the educational climate of the seventies. In this period of rapid change in all educational enterprises, we are foolish not to engage in a little futurology: to analyse trends, predict probable future developments, adjust psychologically to what the future may bring, and plan to meet future needs.

Whether we like it or not, in any foreign or second-language teaching we are swept by the winds of change in general education, in attitudes toward learning, and in community needs.

Paper read at ATESOL First International Conference, Dublin, Ireland, June 29, 1973

FL 004 924

But what has this to do with testing? As Pilliner has stated so succinctly: "It is axiomatic that (the) content [of tests] inevitably influences the teaching and learning which precede them...Properly constructed, (the test) can foster and reinforce good teaching and sound learning and discourage their opposites. To achieve these ends, the test constructor must start with a clear conception of the aims and purposes of the area of learning to which his test is relevant."[1] Reversing Pilliner's emphases, we may say that aims and purposes, construction, and content of tests must be congruent with the aspirations and learning approaches of the day and age. As test constructors, whether for large groups of students from a variety of instructional situations or merely for tomorrow's class, we must not allow ourselves to become so bogged down in the peculiar technical problems of test design that we cannot see the wood for the trees, thus exerting, perhaps involuntarily, a retarding influence on the evolution of foreign-language instruction.

Why do we test anyway? It is salutary sometimes to go back and ask ourselves a question of this type. Is it because it has always been done--because it seems to us of the same order of necessity as the rising and setting of the sun? Is testing an essential part of the learning process? Can we class it as a natural activity? Here we may think of the feedback loop which Miller, Galanter, and Pribram proposed as a

model of the molecular unit of human behavior. Miller and his colleagues called their model TOTE (Test - Operate - Test - Exit).[2] "In its weakest form," they state, "the TOTE asserts simply that the operations an organism performs are constantly guided by the outcomes of various tests."[3] The organism's capability is tested against an existing pattern or criterion. "Action is initiated by an 'incongruity' between the state of the organism and the state that is being tested for, and the action persists until the incongruity...is removed."[4] This criterion having been matched, the organism moves into the next phase of its activity and is challenged to a new effort by the criterion of that phase.

Here we recognize immediately the principle of the thermostat and we can derive an interesting analogy from it for our students' learning. In this model, the test acts as a plan which controls operations. The test phase "involves the specification of whatever knowledge is necessary for the comparison that is to be made, and the operational phase represents what the organism does about it."[5] The test is, as it were, a source of information or a set of instructions which enables the learner to keep up his efforts till he has matched the criterion, testing and retesting to see how close he is coming to the desired performance. Each time he falls short he makes a further effort to reach the criterion; each time he achieves his aim he moves on to the next phase of activity. In this way

the test is an integral part of the learning process: a natural step in any advance. How different this is from the old concept of the test as a hurdle to be surmounted--a hurdle which becomes a discouraging barrier to too many language learners. Here, instead, matching against the criterion, becomes a challenge and a guide to further effort. What a gain it would be if we could convey this attitude toward the test to our students: if they no longer feared the test as a threat to their ego, but saw it as an indicator: a sign on the way. In this paradigm, tests are no longer a special activity, set apart from all others and loaded with unique significance. If we think about it, we can recognize here the distinction between the norm-referenced test, where one student's performance is compared with that of other students or matched against some artificial, external standard, and the criterion-referenced test, where the student knows exactly what knowledge he must demonstrate and either demonstrates it and moves on or cannot demonstrate it and goes back to see how he can improve his performance. This is a revolution in the concept of testing. The student is now responsible for his own learning. When he feels ready to match the criterion, he tests. Note that he tests: it is not we who test him. If necessary, he later retests. When satisfied that he has matched the criterion, he moves on. Idealistic? Perhaps, but it is in keeping with the changing climate in student-teacher relations, in which the student and his needs are central and in which

every student has a right to the opportunity to learn as much as he is able, and as much as he is willing to devote effort to learning. Once the test ceases to be a separate activity, but is interwoven with learning, it may well lose the appearance of what we conceive of as a conventional test.

This being the case, what happens to our "standards"? How do we select those with the level of attainment necessary for specific tasks? This is another question and not an educational one. The educational question must be phrased quite differently! How do we provide the opportunity for each student to attain the highest degree of mastery of a foreign language consistent with his language aptitude, his willingness to devote time and energy, his perseverance, and his interest in the various aspects of language skill and the possible uses of language. A test which is a sorting process for some purpose other than the educational one should properly be assigned to some agency outside of the instructional process. It becomes an admissions procedure, related to job specifications or future study demands. It is not, then, an indicator of achievement or even proficiency, but of what some particular agency is looking for. One agency may need only persons who can read instructions accompanying Japanese stereo equipment; another may wish to accept only students able to read with ease French poetry or experiments in nuclear physics published in English. The person wanting this job, or this form of higher study, will prepare

himself to cope with the particular situation; others will not be judged on their inability to do so. The form of the test will again be a set of instructions for the candidate; the appointment or admission he seeks will be an incentive; the student will attempt to bring his language performance into congruence with the set of instructions and thus exit into the specialized realm he is seeking to enter. As Rebecca Valette has suggested, this may mean the setting of national standards--levels for which students can be certified as required. In English in the U.S.A. Foreign Service Institute levels are often used.

In this case, the source of the set of instructions is clear. At the general level of skill mastery, the source is not so clear by any means. In the past the test was set by the teacher, guided by the syllabus and ultimately by the officials of the system; or it was set by the teacher "alone," guided usually by the textbook he or she had selected (and, therefore, indirectly by certain contemporary "trends", currents, or emphases of the profession, strained through the prejudices and preconceptions of the textbook writer). If the test is indeed to be a set of instructions against which the student may test his developing skills and knowledge, then it must have some clear relationship to his aims and purposes. For this, there must be clear input from the student. There are no "aims of foreign-language instruction". There are only aims of particular students learning a specific foreign language at a particular time

and place. This is the age in which the course must be tailored to the person, not the image.[6] It is at this point that the expertise of the instructor comes into play--in assessing in consultation with the student his special needs, then writing the sets of instructions or tests in relation to these needs. In practical terms this may mean a recipe from the student which the teacher transforms into a suitable test.

We may at this point look closely at two directions in which emphasis on the student as learner has led us and see what implications these emphases have for testing. Apart from a few general suggestions I shall have to leave it to our testing experts to decide what we can do about it. (Some of them are working on it already.) These trends are, however. clearly with us, so we must face their implications squarely. First, each student is an individual with his own preferences as to modality and pace of learning and course content. Secondly, each student is an individual with a personality to express.

The Student as Individual

One does not need to open more than one professional journal to meet the words "individualization of instruction", yet the more one reads the more confused one becomes as to just what this concept means in actual practice. To some it is a new term for an old concept: self-paced instruction through the use of programmed texts--an activity which in practice draws extremely close to independent study. Independent study for

all types of students? Is this providing for individual differences in learning styles and modality and content preferences, or merely for speed of assimilation of what is assigned for learning? Many students, as we well know, do not have the self-confidence to be autonomous learners. This independence can be developed for some with sympathetic encouragement, but meanwhile they prefer to work with others or (is this heresy?) with an instructor. Much so-called individualization is as autocratic in concept as the most rigid teacher-directed classroom, allowing no place for the student to choose other than the independent study of a specific learning packet which his teacher has decreed he shall relish.

Individualizing instruction "means allowing for the different ways in which students learn and giving the students the opportunity to choose what they want to learn, how they want to learn it, and with whom they want to learn it (independently, with other students in a buddy system, in small groups or large groups, or with a teacher, or program)... Individualized instruction will vary according to age and situation, as well as personality and learning style."[7] As Jeannette Veatch has expressed it: "When human interaction takes place on a level where each respects the other's thoughts and ideas, we are getting close to the secret of truly meeting individual differences."[8]

It is to this last concept that we must look for genuine individualization of instruction. As instructors we must give much thought to the ways our individual students learn and the

things in foreign-language learning which "turn them on." The day of the monolithic method and monolithic materials is past. Some students prefer to learn audially, others visually--to some the modality is immaterial. Some learn through abstraction, some through practical, concrete use; some from a book, some from other people; some through a logical progression, some heuristically and, as it seems, almost intuitively. Not only are students different, but so are teachers, and what one can do with ease and composure is difficult for another. There is, then, a move in the selection of materials and techniques toward eclecticism and pragmatism: the teacher seeking what works for him and his present group of students rather than looking for answers from "experts". The "student liberation" movement means "teacher liberation" as well. This trend is to be welcomed in that it restores to the teacher his professional status, requiring of him that he inform himself as fully as he is able so that he can make decisions which are appropriate to the here and now--interpreting students' needs and interests and incorporating them into learning experiences, which in the context of this paper involve interrelated testing. Genuine individualization leads to diversification of approach and content.

Do students really want or need such diversification or is it a current fad? A 1973 study of 1821 undergraduates at the University of Illinois showed that 63 per cent wanted more listening and speaking while 62 per cent wanted more reading.[9]

If these are non-parallel distributions representing distinct preferences for many, as seems plausible, then there is a fair overlap of students, about 25 per cent, who would like more of both, that is, courses which permit them to develop simultaneously in both aural-oral and graphic control of the language, while about 37 per cent at the extremes prefer more exclusively oral communication or reading-oriented courses. The percentages may not be typical for students in foreign-language courses everywhere, but some modification of this distribution seems intuitively to reflect the spread of modality preferences in the community. Content preferences were just as clearly indicated, ranging from practice in communication, learning about contemporary life styles and values, reading of literature, reading of contemporary newspapers and magazines, learning to follow foreign films with ease, translating and interpreting, drawing of information from scientific and technical material, art history, music, cultural and political history, philosophy, and even linguistic analysis. There is no dearth of possible subject matter for any foreign-language learning class. With careful thought and organization much more individualization of content is possible through group work or independent research as students prefer. Groups of students vary. Sometimes we encounter a group all the members of which want to do the same thing; sometimes they divide off visibly. Fortunately diversification will rarely mean a different choice for every student in the group.

Surely such diversification implies real difficulties for testing? If we believe that students must demonstrate the level of proficiency they have reached in comparison with others, then certainly diversity of course content is a complication. If we accept my earlier premise that the test should be an individual matching against a criterion, this need not be so. Instead of better comparative tests, we would seek for a better system of establishing criteria consistent with the student's personal aims and purposes, a system in which the student himself would be involved, so that his progress would be clear, at any point, and continuous. The test as an extra activity then becomes largely superfluous, unless students request it for the pleasure of demonstrating their achievement. When students enjoy tests as a challenge and an opportunity for displaying what they know, we shall have reached the optimal form and timing of the test. If grades are necessary (and we may question whether they are in most cases), then the grade should be based on what the student has achieved as an individual; they should reflect his personal effort and progress toward an individual goal.

Once the one standard test becomes the goal of all students, we are back where we started. Because of the limitations of the job market or of college entrance, or because some other need for an elite has arisen, it becomes depressingly inevitable that some of the most hardworking students will fail. This is built into the system. Is such artificial "failure" necessary or desirable? If not, let us change the system.

The Student as an Individual with a Personality to Express

Thus far, let's admit it, even our discussion of providing for the student as an individual has still been in the main teacher-directed, with the teacher making decisions about what is best for the student. Of course, it has involved the teacher taking the student into his confidence: making sure he understands what he is going to do and why, just what he will be expected to demonstrate and how, what criterion he must match before he moves on. This is built into the concept, for instance, of performance objectives. In the words of a recent book: "The teacher...decides in advance which features of the unit he intends to stress in his classes and what degree of proficiency he wants the students to develop with respect to those features...It is up to the teacher to set the level of mastery, but his intention should always be that as many students as possible attain a high score."[10] If the student is a full partner, he must also have the opportunity to tell the teacher what his expectations for the course are, what particular skills and course content interest him, when instruction is moving too fast or too slow for him, and, at a particular moment, the specific aspects of his study on which he feels the need to test himself.

In theory, performance objectives seem a good idea: they enable teacher and student to come to a clear understanding of the next step to be surmounted. In practice, in foreign-language

learning they can be very confining. A foreign language to be learned cannot be reduced to a multitude of small elements which we accumulate like beads on a string (this phonological discrimination, that use of the past tense in indirect speech, ten words for parts of the body), and who is to say that in each case the student must know these to the stage where he makes only one spelling mistake in ten examples of their use, or fails to make some distinction in only two cases out of twelve?

For some years now, leaders in our field have been pointing out that use of a foreign language is more than the sum of its parts, that there is macro-language use as opposed to micro-language learning. The micro approach can stultify foreign -language learning even in its early stages. Naturally, if the student is to use language he needs a basic knowledge of phonology, grammar, and lexicon, but these must continually be practiced in some form of real production of meaningful messages. In other words, the student as individual learner must have the opportunity to express himself through the language in terms of his own personality, in some use of the language for the natural purposes of language: as part of an interaction of communication, either giving or receiving, in speech or in writing.[11] If this interchange is a natural expression of personality it cannot be predetermined with an established criterion level of mastery of the nine out of ten variety: the criteria in

these cases are comprehension and comprehensibility, qualities which are very difficult to quantify.[12]

In all education (shall we say in all living) we have this continual tension between man's desire to organize, to bring order to phenomena, to quantify, and unfettered natural growth: between the classical and the romantic impulse, between control and self-expression, between the Robert Hutchins and the Ivan Illichs. As educators we have to keep our balance between the two as the pendulum swings. Performance objectives seem to bring order, clarity, direction, rational progression to foreign-language learning, yet, given preeminence, they stunt the fragile plant they are there to nurture, just as too rigorous pruning and training may produce an espalier but not a free-standing tree in a natural garden.

Here I shall return to one of my favorite themes since 1964,[13] the fact that in language learning we have to control language at two levels. There is basic core learning of the phonological, morphological, and syntactical operations of the language and of the interrelationships of these systems with the semantic system. This is what I am referring to as micro-language learning. Mastering it is essential, time-consuming, sometimes tedious, hard work and it is here that the performance objective approach and its related modes of testing are useful. It is the second level, or macro-level, of the natural use of the new language for the expression of personal

meaning which we seem continually to neglect, but which is absolutely essential if the learning of the language is not to be time-wasting busywork. This macro-language use is not a later advanced stage of study which we are sorry that many of our students do not reach: it is our major purpose in foreign-language instruction which must be encouraged and fostered from the first elementary learnings. It is this level of language use which cannot be confined by the conventional performance objectives if it is to retain that spontaneity which is its hallmark. If we wish to encourage creativity and self-expression in the use of the second language, we cannot decide in advance what features our students will use and the degree of proficiency we want them to demonstrate in the use of these features.

How, then, can this spontaneous language use be tested? Valette says: "Until we know precisely what we intend to teach we cannot measure our success."[14] Perhaps with natural use of language we will never know <u>precisely</u> what we must teach, but yet it seems clear that the natural use of language by our students is more important than a clear-cut "measure" of our success. We cannot teach for creativeness in language use with functional comprehension and comprehensibility as the ultimate criteria and then test for mere accuracy of detail, as most standard tests seem to do at present.

It is here that we need better tests than the standard interview for speaking and free composition for writing. Var-

ious possibilities have been suggested involving our giving the student a real task to perform which requires him to seek information and convey information in the language (in speech or writing as his needs require) and then evaluating him on the successful completion of the task. In concrete terms, this suggestion has endless possibilities for actualization at various levels of difficulty and can be adapted to specific uses of language which interest a particular student (the conducting of business affairs, the enjoyment of a film or play, the investigation of a scientific problem). At an elementary stage, the test can entail approaching a monolingual or presumed monolingual speaker of the language to find out such information as his name, age, address, telephone number, and occupation for entering on a file card. More testing along these lines would make the test a natural and enjoyable part of the learning, as was proposed in the earlier part of this paper.[15] It would also provide a climax to a unit of study which the student could anticipate with pleasure as an opportunity to test himself against a criterion of authentic communication. This type of testing requires imagination and ingenuity on the part of the examiner, and our testing experts could perform a service for the profession in drafting and publishing a number of tests along these lines, with suggested adaptations to keep them varied. Security would not be a problem in these cases because each actualization of the test would take a different turn as

the situation was followed through by the student being tested. The result of the test would be rated either successful or unsuccessful (either the student was able to carry the task through to a satisfactory conclusion or he was not), so that the "subjective" element in the judgment of the examiner would not be of any great significance. Clearly there would be variability in the amount and complexity of the communication which took place, but the student would have demonstrated his ability to give and receive information or to interact informally in an acceptable manner.

Carroll tells us that "from a practical point of view it may often suffice to construct tests that measure only integrated performance based on competence. For example, a general test of proficiency in a foreign language is often found to yield just as good validity when its items are complex, each drawing upon a wide sample of linguistic dompetences, as when each item has been contrived to tap competence in one and only one specific feature of the foreign language...apparently the extent to which a language test should attempt to measure specific aspects of competence depends upon its purpose--that is, the extent to which there is need for diagnosis of specific skills as opposed to a generalized, overall assessment of proficiency."[16] If, as Pilliner says, the form of the test gives direction to learning, then it is essential that it be consonant with the aim of natural language use if it is to be valid in the contemporary context. Validity is a much-prized con-

cept in testing, yet too many tests are still based on the aims of a decade or two decades ago, lagging behind materials and classroom instruction. Thus they retard the evolution of a progressive view of language teaching, instead of clarifying goals for the less informed.

With all global testing, the perennial problem arises: how much accuracy in detail should we expect or require if effective communication is to be the goal? The answer to this must realistically be relative: the business man out for a contract cannot afford to misunderstand detail or to give assurances which can be misinterpreted. The scientist writing a research paper must state exactly what he intends to state if the equipment is not to blow up when the experiment is replicated. These people need a degree of accuracy not usually essential for the tourist or the captain's wife on a foreign base. Such students would, by their training, realize the need for accurate expression in professional matters and presumably be motivated to work toward it. The businessman, more than the scientific researcher, would understand the need for accuracy in intonational patterns and pitch levels so that he would not sound angry when he was intending to be persuasive. On the other hand, many an emissary in a foreign country has found that a certain degree of foreignness in his speech patterns, far from being a handicap, elicited a greater tolerance on the part of local people toward his early mistakes in adapting to the cultural patterns of their society.

More research is needed in all languages into those elements for which an absolute degree of accuracy is required for communication which will not irritate the native speaker, and those which the native speaker will accept as amusing but pleasant indications that the stranger has really tried to learn his language and meet him on his own ground. What we must remember is that we can so easily kill, or at least considerably dampen, the enthusiasm of a foreign-language learner by preferring accuracy of detail to sincere efforts to create spontaneous utterances or write expressive prose. When it comes to micro-testing, we may also remember that the person capable of macro-performance may well be able to cope with the details of the micro-test, but that each person passing the micro-test is not necessarily able to perform acceptably at the macro-level. The fact that the micro-test is so much easier to administer is a danger to us as a profession. If we become addicted to fill-in-the-blank tests and multiple-choice items, we must not be surprised if our students think that this is what performance in a foreign language really is. Let us remember that by our testing they shall know us, far better than we shall know them.

Notes

1. A.E. Pilliner, "Subjective and Objective Testing" in Alan Davies, ed., Language Testing Symposium. A Psycholinguistic Approach. (O.U.P., 1968, p.31).

2. G.A. Miller, E. Galanter, and K.H. Pribram, Plans and the Structure of Behavior (Holt, Rinehart and Winston, 1960). The TOTE unit is described and discussed on pp.25-38.

3. Ibid., p.29.

4. Ibid., pp.25-26.

5. Ibid., p.31.

6. This subject is developed in relation to undergraduate level courses in Wilga M. Rivers, "The Non-Major: Tailoring the Course to the Person--not the Image" in K. Jankowsky, ed., Language and International Studies (Georgetown University Press, Monograph No.26), forthcoming. Co-publication: ADFL Bulletin, Vol.4, No.2 (Dec., 1973).

7. Wilga M. Rivers, "Techniques for Developing Proficiency in the Spoken Language in an Individualized Foreign Language Program," in H.B.Altman and R.L. Politzer, eds., Individualizing Foreign Language Instruction (Rowley, Mass: Newbury House, 1971), p.165.

8. "Individualizing," in Virgil M. Howes, ed., Individualization of Instruction. A Teaching Strategy (New York: MacMillan, 1970), pp.91-2.

9. Reported in Wilga M. Rivers, "The Non-Major..." (forthcoming).

10. E.D. Allen and R.M. Valette, Modern Language Classroom Techniques (New York: Harcourt Brace Jovanovich, 1972), pp.21, 37.

11. I have discussed how we may move in this direction from the earliest stages in two papers: "Talking off the Tops of their Heads" in W.M. Rivers, Speaking in Many Tongues (Rowley, Mass.: Newbury House, 1972), and "From Linguistic Competence to Communicative Competence" in TESOL Quarterly, 7, 1(1973), 25-34.

12. In R.M. Valette and R.S. Disick, Modern Language Performance Objectives and Individualization (New York: Harcourt Brace Jovanovich, 1972), the distinction is made between "formal performance objectives" and "expressive performance objectives," the latter being "open-ended" with the conditions "less precise" (p.26), "the actual form such student behavior will assume (being) not as readily predictable" (p.54).

13. See Wilga M. Rivers, The Psychologist and the Foreign-Language Teacher (Chicago: University of Chicago Press, 1964), Chapter 6: "Two Levels of Language."

14. R.M. Valette, Directions in Foreign Language Testing (New York: MLA/ERIC, 1969), p.31?

15. Many of the activities described in "Talking off the Tops of their Heads" in Rivers (1972) could be used as tests of this type.

16. J.C. Carroll, "The Psychology of Language Testing" in Alan Davies, ed., Language Testing Symposium (O.U.P., 1968), pp.56-58.